TITLE PAGE

FINAL REPORT

INSTITUTION: The University of Wisconsin-Madison

PRINCIPAL INVESTIGATOR: Michelle R. Rondon

ADDRESS: Department of Bacteriology
Room 151 Biochemistry
420 Henry Mall
Madison Wisconsin 53706

TELEPHONE NUMBERS

Office (608) 265 7767
Department (608) 262 2914
Fax (608) 262 9865

E-MAIL ADDRESS: mrrondon@wisc.edu

TITLE: Global regulatory pathways in the alpha-proteobacteria

DOE AWARD NUMBER: DE-FG02-02ER15343

Michelle R. Rondon
The University of Wisconsin-Madison
DOE AWARD NUMBER DE-FG02-02ER15343

Project Abstract

A major goal for microbiologists in the twenty-first century is to develop an understanding of the microbial cell in all its complexity. In addition to understanding the function of individual gene products we need to focus on how the cell regulates gene expression at a global level to respond to different environmental parameters. Development of genomic technologies such as complete genome sequencing, proteomics, and global comparisons of mRNA expression patterns allows us to begin to address this issue.

This proposal focuses on a number of phylogenetically related bacteria that are involved in environmentally important processes such as carbon sequestration and bioremediation. Genome sequencing projects of a number of these bacteria have revealed the presence of a small family of regulatory genes found thus far only in the alpha-proteobacteria. These genes encode proteins that are related to the global regulatory protein RosR in *Rhizobium etli*, which is involved in determining nodulation competitiveness in this bacterium.

Our goal is to examine the function of the proteins encoded by this gene family in several of the bacteria containing homologs to RosR. We will construct gene disruption mutations in a number of these bacteria and characterize the resulting mutant strains using two-dimensional gel electrophoresis and genetic and biochemical techniques. We will thus determine if the other proteins also function as global regulators of gene expression.

Using proteomics methods we will identify the specific proteins whose expression varies depending on the presence or absence of the RosR homolog. Over fifty loci regulated by RosR have been identified in *R. etli* using transposon mutagenesis; this will serve as out benchmark to which we will compare the other regulons.

We expect to identify genes regulated by RosR homologs in several bacterial species, including, but not limited to *Rhodopseudomonas palustris* and *Sphingomonas aromaticivorans*. In this way we will provide valuable information on gene regulation in this group of bacteria, expand our understanding of the evolution of global regulatory pathways, and develop methods for comparative regulon analysis among microbes.

Michelle R. Rondon
The University of Wisconsin-Madison
DOE AWARD NUMBER DE-FG02-02ER15343

STATEMENT OF RESEARCH PROBLEM

Whole genome sequencing has given unprecedented insight into the physiological capabilities of microorganisms. Understanding how this genetic information is coordinately regulated to generate an environmentally responsive cell is essential for a complete understanding of genome function. This question is even more fascinating when new whole-genome sequences uncover the possibility of a novel global regulatory system shared between, and limited to, a group of phylogenetically related microbes. Recent genome sequencing projects of several alpha-proteobacteria have uncovered a small but intriguing family of open reading frames (ORFs). These are related to *rosR* from *Rhizobium etli*, which encodes a zinc-finger-containing transcription factor that is a global regulatory protein involved in determining nodulation competitiveness in this bacterium. These ORFs are conserved over their entire length, yet some lack certain amino acid residues implicated in zinc finger formation, which was shown to be important for activity of the related Ros protein of *Agrobacterium tumefaciens*. The presence of these homologs raises several interesting questions. First, what functional predictions can be made about ORFs that lack conserved amino acid residues? Second, what predictions can be made about the conservation of regulatory circuitry based on finding a homolog to a given transcription factor in a genomic sequence? Third, given that RosR in *R. etli* regulates a large number of genes and is important for determining competitive fitness of this bacterium, what is the role of the new homologs in the biology of that bacteria that contain them?

To address these questions, we will implement molecular and genomic approaches to investigate the function of this gene family and to identify genes that might be regulated by the encoded proteins. This project is of interest to the mission of the Department of Energy since the RosR-related ORFs have been found mainly within genomes of bacteria that are DOE-relevant. These are *Sphingomonas aromaticivorans* (deep subsurface, bioremediation), Rhodopseudomonas palustris (biodegradation, carbon management), *Caulobacter crescentus* (bioremediation), *Methylobacterium extorquens* (carbon management), and *Magnetospirillum magnetotacticum* and *Magnetococcus* MC-1 (magnetosomes). Besides their roles in these important environmental problems, many aspects of the biology of these bacteria are for the most part relatively poorly understood. The research described here will provide new information on gene regulation in this group of bacteria.

RESEARCH OBJECTIVES

1. Genetic analysis of RosR homologs. Our initial approach to understand the function of these proteins will be to construct gene disruption mutations in several of the homologs, and analyze the phenotypes of the resulting strains.

2. Biochemical characterization of RosR homologs. Our long-term goal is to understand the cellular role of this gene family. Since RosR is better characterized in terms of its target genes, we will continue our investigation of the function of RosR, and compare it to the newly identified Ros homologs.

3. Comparative proteome analysis to identify genes regulated by the RosR homologs. We will use proteomics to analyze proteins expressed in the wild type and mutant strains of bacteria that show different protein expression patterns in Objective 1.

RESEARCH RESULTS

1. Analysis of the RosR homolog in *Rhodopseudomonas palustris*.

Rhodopseudomonas palustris contains one homolog to *rosR*, called RPA1852 (6, previously called *rpal3127)*. We constructed mutations in RPA1852 by inserting a cassette containing a *gusA* reporter gene (encoding β-glucuronidase) into the RPA1852 gene cloned onto a suicide plasmid. These mutations were introduced into *R. palustris* and mutant strains were identified that had exchanged the mutant version of the gene for the wild-type version. The correct genomic structure was confirmed by Southern blot analysis. We constructed a mutant strain containing the promoterless reporter gene in the same orientation as RPA1852 and a control mutant strain containing the promoterless reporter gene in the opposite orientation as RPA1852.

There was no observable colony phenotype associated with the mutant strains, and EPS and LPS analysis showed no differences between the wild type and the mutant strains (data not shown).

RPA1852 was expressed in *R. palustris*, based on the results of quantitative GUS assays. As shown in Table 1, we assayed β-glucuronidase expression in wild type *R. palustris* and in the two mutants in RPA1852.

Table 1. Expression of RPA1852.

Strain tested	Activity in early log phase	Activity in late log phase
R. palustris	<1	<1
RPA1852::*gusA*1	200	100
RPA1852::*gusA*2	10	7

Activity refers to β-glucuronidase activity measured in permeabilized cells. RPA1852::*gusA*1 contains a gusA fusion in the same orientation as the RPA1852 gene, while RPA1852::*gusA*2 contains a gusA fusion in the opposite orientation to RPA1852.

We tested the ability of RPA1852 to regulate expression of its own gene, but introduction of RPA1852 on pHRP309 (a replicating plasmid in *R. palustris*) had no effect on expression of RPA1852::*gusA*1 (data not shown).

Total soluble protein from the wild type and mutant strains was subjected to two-dimensional protein gel analysis to identify proteins that were differentially expressed in one strain versus the other. Unfortunately, despite multiple experiments, no proteins were identified that varied between the strains (data not shown). These gels were run by Kendrick Laboratories, a service located in Madison, WI that performs two-

dimensional gel electrophoresis and other proteomic analyses (http://www.kendricklabs.com).

2. Analysis of the RosR homologs in *Rhodospirillum rubrum*.

R. rubrum contains two genes homologous to *rosR*, Rru_A3766 and Rru_A1829 (genome is available at GenBank under the submission number NC_007643). We constructed mutations similar to those described above in Rru_A3766. Mutants were confirmed by PCR analysis. Again, there was no observable colony phenotype associated with the mutant strains. The gene was expressed in *R. rubrum*, as measured by β-glucuronidase activity. It was difficult to obtain repeatable b-glucuronidase activity data. The mutant strain with the *gusA* insertion in the same orientation as Rru_A3766 consistently gave a high level of β-glucuronidase activity (between 90 and 680 units of activity) whereas the mutant strain with the gusA insertion in the opposite orientation as Rru_A3766 consistently gave very low activity (between 5 and 8 units). Again we tried complementation of the mutation by providing Rru_A3766 on pRK310, a replicating plasmid in *R. rubrum*, however strains containing this plasmid showed no changes in β-glucuronidase activity (data not shown).

Total soluble protein from the wild-type and mutant strains was subjected to two-dimensional protein gel analysis, as described above, to identify proteins that were differentially expressed in one strain versus the other, but unfortunately, no proteins were identified (data not shown). These gels were also run by Kendrick Laboratories, a service located in Madison, WI that performs two-dimensional gel electrophoresis and other proteomic analyses (http://www.kendricklabs.com).

We noted that Rru_A3766 is located immediately upstream of a gene encoding a homolog to the global iron regulatory protein Fur. We used reverse transcriptase PCR to show that the genes were co-transcribed, and that the PCR product representing the two-gene transcript was not produced when RNA from the Rru_A3766 mutant strain was used as a template. This would be the first evidence for a role for Rru_A3766 in the cell.

3. Characterization of RosR in *R. etli*.

3A. Analysis of RosR-regulated promoters.

We are interested in defining the DNA site to which RosR binds, as a first step towards analysis of RosR function. Prior to publication of the *R.etli* genome sequence, we sequenced the surrounding regions of several RosR-regulated loci, to identify likely promoter regions. These regions were then cloned into the gusA promoter-probe vector pFUS1 (7). The resulting clones were mobilized into the wild-type *R. etli* and into a *rosR* mutant derivative, and quantitative GUS assays were performed. As shown below, we identified several RosR-regulated promoter fragments.

Table 2. Identification of RosR-regulated promoter fragments. Activity refers to GUS activity of the reporter plasmid.

Mutant designation	R. etli genome designation	Activity in wild type strain	Activity in *rosR* mutant strain	Fold regulation	Size of insert (bp)
MB01	RHE_PE00197	140	7790	56	661
MB02	RHE_CH03393	75	2580	35	447
MB48	RHE_CH03346	355	1485	4	575
RosR	RHE_CH01249	1285	10000	8	795
pFUS1	N.A.	49	49	N.A.	None

Promoters are named based on the original mutant designation of Bittinger and Handelsman (3). Genome designation follows the *R. etli* genome project (5).

It is possible that these promoters are only indirectly regulated by RosR, and are directly regulated by some other transcription factor that is itself regulated by RosR. This may be a likely even since several regulatory genes were identified in a search for RosR-regulated genes (3). We made multiple attempts to develop a two-plasmid system in *E. coli*, using one plasmid to express RosR and the other plasmid to contain a RosR-regulated promoter controlling a reporter gene, but our attempts at this failed. We concluded in this case the RosR itself was affecting the reporter plasmid independently of the promoter region cloned into it and was thus not an accurate measurement of RosR regulation (data not shown).

3B. Analysis of a locus involved in nodulation competitiveness.

In a parallel investigation to that described in section 3A, we analyzed regulation of two loci in *R. etli* that are regulated by RosR. These were mutants previously described as affecting nodulation competitiveness in a RosR-dependent manner (3). Further sequencing of the region revealed that one mutant, MB 13, was disrupted in a putative transcription factor with similarity to transcription factors in the CRP/FNR family. The other mutant, MB65, was disrupted in a locus encoding a putative exopolysaccharide production protein, and was transcribed divergently from the gene disrupted in MB13. The R. etli genome sequencing project refers to these loci as RHE_CH03384 (MB13) and RHE_CH03385 (MB65).

We hypothesized that the transcription factor encoded by RHE_CH03384 was likely directly regulated by RosR, and it in turn directly regulated RHE_CH03385. To test this, we constructed a mutation in RHE_CH03384, using a kanamycin resistance cassette (7). We also cloned promoter regions from the two loci into pFUS1. The resulting plasmids were introduced into the different strains and GUS activity was measured.

Table 3. Activity of RHE_CH03384 and RHE_CH03385 promoter regions in different genetic backgrounds. Activity refers to β-glucuronidase (GUS) activity of the reporter plasmid.

Mutant designation	*R. etli* genome designation	Activity in wild type strain	Activity in *rosR* mutant strain	Activity in RHE_CH03384 mutant strain	Activity in *rosR* RHE_CH03384 mutant strain
MB13	RHE_CH03384	110	10390	105	15295
MB65	RHE_CH03385	35	11905	20	155

Promoters are named based on the original mutant designation of Bittinger and Handelsman (3). Genome designation follows the *R. etli* genome project (5).

These results support our hypothesis that these loci form a regulatory cascade; it is likely that RosR directly regulates RHE_CH03384, which in turn is a specific transcription factor for RHE_CH03385. RHE_CH03384 is needed to activate expression of RHE_CH03385, since the strain containing the mutation in RHE_CH03384 fails to activate expression of RHE_CH03385, even in the *rosR* mutant background.

We also noted that mutations in these loci affected the ability of *R. etli* to attach to solid surfaces. Whereas the wild type strain did not attach to plastic microtiter plates, the *rosR* mutant formed a tight layer that would not be washed out of the wells with water. The mutation in RHE_CH03384 abolished the ability of the cells to adhere. This suggests that RHE_CH03384 controls production of a polysaccharide that may affect the ability of *R. etli* to adhere to surfaces. The locus containing RHE_CH03385 is a candidate for this polysaccharide biosynthesis locus.

3C. Analysis of the RosR promoter

RosR had been previously shown to regulate its own expression (2). We also constructed plasmid derivates of pFUS1 containing fragments of the RosR promoter and tested RosR regulation of these plasmids. We found that a fragment of 795 base pairs including to potential *rosR* promoter region was 8 fold regulated by RosR (10,000 units of activity in the *rosR* mutant strain and 1285 units of activity in the wild type strain).

We used RT-PCR to further address this issue. Using primers at varying distances from the *rosR* translational start site, we found that the *rosR* gene likely contains a 5' untranslated region of between 265 and 296 nucleotides. This is similar to previous results seen with the *mucR* gene (the *rosR* homolog in *Sinorhizobium meliloti*), where an upstream untranslated region of 425 nucleotides was suggested (1).

3D. Analysis of RosR protein activity.

Previous researchers had demonstrated *in vitro* binding activity of RosR (4). We were unable to repeat that activity, despite repeated attempts. We produced RosR as a hybrid protein to the Maltose Binding Protein, in an attempt to improve solubility of the protein. Preliminary work with purified MBP-RosR suggested the protein may form a hexamer, but we were not able to show *in vitro* binding activity with this protein. We also produced RosR as a fusion protein with Glutathione S-Transferase, but again, this protein showed no *in vitro* activity. We concluded that production of active RosR was not reliable, despite the published reports.

References

1. P. A. Bertram-Drogatz, S. Rüberg, A. Becker, A. Pühler. 1997. The regulatory protein MucR binds to a short DNA region located upstream of the *mucR* coding region in *Rhizobium meliloti*. Molec. Gen. Genet. 254: 529-538.

2. M. A. Bittinger, J. L. Milner, B. J. Saville, and J. Handelsman. 1997. RosR, a determinant of nodulation competitiveness in *Rhizobium etli*. Molec. Plant-Microbe Interact. 10:180-186

3. M. A. Bittinger and J. Handelsman. 2000. Identification of Genes in the RosR Regulon of *Rhizobium etli*. J. Bacteriol. 182: 1706-1713.

4. M. A. Bittinger, J. A. Gross, J. Widom, J. Clardy, and J. Handelsman. 2000. *Rhizobium etli* CE3 carries *vir* gene homologs on a self-transmissible plasmid. Molec. Plan-Microbe Interact. 13:1019-1021.

5. V. González, R. I. Santamaría, P. Bustos, I. Hernández-González, A. Medrano-Soto, G. Moreno-Hagelsieb, S. Chandra Janga, M. A. Ramírez, V. Jiménez-Jacinto, J. Collado-Vides, and G. Dávila. 2006. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons Proc. Natl Acad Sci U S A. 103: 3834–3839.

6. F. W. Larimer, P. Chain, L. Hauser, J. Lamerdin, S. Malfatti, L. Do, M. L. Land, D. A. Pelletier, J. T. Beatty, A. S. Lang, F. R. Tabita, J. L. Gibson, T. E. Hanson, C. Bobst, J. L. Torres y Torres, C. Peres, F. H. Harrison, J. Gibson & C. S. Harwood. 2004. Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris.* Nature Biotechnol. 22: 55-61.

7. W.G. Reeve, R.P. Tiwari, P.S. Worsley, M.J. Dilworth, A.R. Glenn, and J.G. Howieson. 1999. Constructs for insertional mutagenesis, transcriptional signal localization and gene regulation studies in root nodule and other bacteria. Microbiol. 145: 1307 - 1316.